

THEORY AND OBSERVATION IN FAMILY EPIDEMICS OF THE COMMON COLD

BY

M. A. HEASMAN* AND D. D. REID

Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine

Recent developments in the mathematical theories of chance behaviour have quickened interest in their applications to epidemic experience. Theoretical epidemiology aims at the prediction of the likely flow of infection within a group given certain assumptions about the laws underlying epidemic events. Ross (1915), for example, deduced the rise and fall of malarial incidence which would derive from some basic axioms about the frequency of infective mosquitos biting susceptible hosts during their period of contact; Macdonald (1953) has brought this approach up to date by comparing field observations of malarial epidemics with expectation based on a mathematical model of the likely pattern of events. The practical importance of such work lies in the possibility of distinguishing those theories of epidemic behaviour which best fit the observed facts, in the hope that this will lead to a clearer understanding of the factors which determine that behaviour.

Most of the work in the field has dealt with major epidemics in large populations using "deterministic" models, *i.e.* when the future course of the epidemic is precisely determined by basic assumptions about the numbers of susceptibles and the frequency and intimacy of their contact with an infective source. In smaller groups, the likelihood of one susceptible making effective contact with an infectious member of the same group cannot be exactly predicted, for the element of chance bulks larger as the size of the group decreases. In these circumstances, this chance element has to be more carefully taken into account, and here the newer theories of stochastic behaviour have proved useful. Bailey (1957) gives a scholarly review of this development. It is enough to say that Greenwood (1931) introduced the idea that the spread of measles within a household could be described in terms of a series or chain of chance events which followed the usual binomial expression of the law of probability. Any one of three susceptible

children exposed to an infection in another member of the family may, in general, have a 10 per cent. chance of getting the disease; but the number of times all would escape or one, two, or three be infected will follow the usual laws of chance represented by the expansion of the binomial $(\cdot 10 + \cdot 90)^3$. If only one is infected, the chance distribution of the frequency of the remaining children becoming infected in the next stage of the family epidemic is again represented by the binomial $(\cdot 10 + \cdot 90)^2$. In short, the pattern of a family epidemic follows the dictates of a chain of such binomials. Essentially the same line of argument had already been followed by Reed and Frost (Bailey, 1957) in developing their theory of the epidemic behaviour of measles in larger communities. But there was one important difference. Both theories assumed that the period of infectiousness was relatively brief, that the period of incubation was relatively long and constant in duration, and that the susceptibles, once infected, were effectively removed from circulation. In the Reed-Frost approach, however, the chance of infection among the remaining susceptibles at any particular stage of the epidemic depended on the number of infective subjects circulating at that stage, to which these susceptibles were exposed. Greenwood's model took no account of the possibility that this had any material effect. Lidwell and Somerville (1951) developed a model, algebraically identical to the Reed-Frost conception, which they applied to the results of their field study of colds in families in a village community. They found that their modification of Greenwood's chain binomial accorded better with their observations than did the original.

As Greenwood has remarked, mothers of large families need no refined statistical method to satisfy them that measles is an infectious disease; but the difficulty of producing evidence of infectiousness in less clearly defined maladies is evident in the inability of Hope Simpson and Sutherland (1954) to fit observation to epidemic theory in familial occurrences of

* Present address: General Register Office, Somerset House, W.C.2.

influenza. The need is for observational data of sufficient precision and detail to test the adequacy of the description of biological processes given by different epidemiological theories. An investigation of respiratory illnesses in families already reported (Brimblecombe, Cruickshank, Masters, Reid, and Stewart, 1958) produced the data discussed in the present paper.

SOURCE AND NATURE OF MATERIAL

In this investigation, families of the same size, two parents and three children, living in different circumstances of domestic overcrowding were visited at fortnightly intervals. The date of onset and the clinical nature of upper respiratory infectious experienced by each member of the family were charted on a time scale marked off in days. Family epidemics of acute coryza—or common colds—were thus available for analysis.

By inspection of the epidemic time charts, it was possible to identify new or primary introductions of illness into the household by the onset of a cold after a lapse of 10 days since the last such case in the same home. Two such cases occurring on the same or succeeding days were classified as multiple primaries. Thereafter, the links in the epidemic chain of spread were defined by an interval of one day or more between successive cases in the same family. These family epidemics could then be described thus—1-2-1, 1-1-1-0, 2-1-0, etc. It must be emphasized that although this method of classification is somewhat arbitrary, it was completed before the corresponding theoretical distributions were worked out and the interval chosen agrees with the distribution of presumptive incubation periods of the common cold seen in field surveys (*e.g.* Badger, Dingle, Feller, Hodges, Jordan, and Rammelkamp, 1953).

RESULTS

As already noted, the first stage of a family epidemic after a new single primary introduction should, in both the Greenwood and the Reed-Frost formulations, follow the simple binomial distribution of chance events. Table I compares the observed distribution of the number of the four susceptibles in these families then affected with the distribution expected on this purely random basis. The fit is excellent.

TABLE I
OBSERVED AND EXPECTED DISTRIBUTION OF COLDS
IN THE FIRST STAGE OF FAMILY EPIDEMICS
(SINGLE PRIMARY ONLY)

No. of Further Cases	No. of Families	
	Observed	Expected
0	423	420.91
1	199	203.18
2	39	36.79
3	3	3.05
4	0	0.07
Total	664	664.00

Probability of Transfer = 0.108
 $\chi^2 = 0.205$; *d.f.* = 1; 0.70 > *P* > 0.50

When this experience is broken down according to the family status of the introducing case, the fit of the observed to theoretical distribution of subsequent cases is obtained in Table II. The probabilities of transfer differ, of course, according to the status of the introducing or index case and the susceptibility of the remainder, and the significance of these differences have already been discussed (Brimblecombe and others, 1958). Although host susceptibility certainly differs between members of the family, these differences are not large enough to affect the fit of the total distribution seen in Table I.

TABLE II
OBSERVED AND EXPECTED DISTRIBUTIONS OF COLDS IN FIRST STAGE OF FAMILY EPIDEMICS
ACCORDING TO THE STATUS OF THE INTRODUCING MEMBER (SINGLE PRIMARY ONLY)

No. of Further Cases	Status of Introducing Case							
	Father		Mother		School Child		Pre-school Child	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
0	53	55.0	75	73.4	148	144.3	147	149.2
1	31	27.4	25	27.5	77	84.3	66	62.3
2	4	5.1	4	3.8	22	18.6	9	9.7
3	0	0.4	1	0.2	2	1.8	0	0.7
4	0	0.0	0	0.0	0	0.1	0	0.0
Total	88	87.9	105	104.9	249	249.1	222	221.9
Probability of Transfer	.111		.086		.128		.095	
χ^2	1.18		3.47		1.47		1.00	
<i>P</i>80 > .70		.50 > .30		.70 > .50		.90 > .80	

Table III shows distribution of the total size (*i.e.* the number of susceptibles attacked) of those family epidemics actually observed compared with the total size to be expected, either on Greenwood's theory which ignores the added risk to susceptibles exposed to more than one infectious case, or on the Reed-Frost model which makes an allowance for this possibility. Clearly, the Reed-Frost theory is in better accord with the observed facts; but neither gives a very satisfactory prediction of the distribution of total epidemic size. When, however, the comparison of observation with expectation on the basis of the Reed-Frost model is made within groups of families living in similar housing conditions, the fit, seen in Table IV, is considerably improved by thus accounting for one source of variation.

TABLE III
OBSERVED AND EXPECTED DISTRIBUTIONS OF COLDS
IN COMPLETED FAMILY EPIDEMICS
(SINGLE PRIMARY ONLY)

No. of Further Cases	No. of Families		
	Observed	Expected	
		Reed and Frost	Greenwood
0	423	409.9	405.3
1	131	146.3	147.1
2	60	68.9	79.3
3	36	24.4	31.0
4	14	14.4	1.2
Total	664	663.9	663.9
		$\chi^2 = 8.69$ $n = 3$ $.05 > P > .02$	$\chi^2 = 17.07$ $n = 2$ $P = < .001$

TABLE IV
OBSERVED AND EXPECTED DISTRIBUTION ON THE
REED-FROST MODEL ACCORDING TO DEGREES
OF OVERCROWDING

Further Cases	Domestic Overcrowding					
	Overcrowding		Crowded		Uncrowded	
	Ob- served	Ex- pected	Ob- served	Ex- pected	Ob- served	Ex- pected
0	112	106.8	155	145.7	156	158.1
1	35	40.5	41	53.6	55	51.6
2	17	20.7	24	26.3	19	21.8
3	11	8.3	15	9.9	10	6.4
4	6	4.6	6	5.4	2	4.1
Total	181	180.9	241	240.9	242	242.0
Probability of Transfer	$\chi^2 = 2.97$ $n = 3$ $.70 > P > .50$		$\chi^2 = 6.45$ $n = 3$ $.10 > P > .05$		$\chi^2 = 3.71$ $n = 3$ $.30 > P > .20$	
	.124		.118		.101	

In Table V the family epidemics are classified into chains of different types, which include multiple as well as single primary cases. The observed frequencies

agree well with the expectation based on the Reed-Frost model. For the single primary chains, the probability of infection is calculated from all the generations of the epidemic in contrast to the probability used in Table I, which is derived from first generation cases only. The probability of infection calculated in the double primary chains is derived in the same way from all generations, but it is assumed that secondary cases are infected by only one source. Thus, in the first generation, half of the cases occurring are attributed to one primary and half to the other. In estimating the probability of transfer between any two members of the family, the number of cases resulting from exposure to two sources of infection in the previous generation is therefore reduced by half. Calculated in this way, the probability of transfer during double primary epidemics is seen in Table V to be less than in outbreaks following a single introduction. This may be due to the simultaneous infection of two of the more susceptible members of the family at some common source such as the school, and a corresponding reduction in the transfer rate among the remaining more resistant members. Alternatively, the erroneous classification of an early secondary as a primary will lower the apparent risk of infection among the others in the household. The good fit between observed and theoretical values makes the latter possibility seem unlikely.

TABLE V
OBSERVED AND EXPECTED FREQUENCIES OF SINGLE
OUTBREAKS ACCORDING TO REED-FROST MODEL

Chain	Observed	Expected	Chain	Observed	Expected
1-0	423	409.9	2-0	55	55.5
1-1-0	131	146.2	2-1-0	27	25.8
1-2-0	24	24.8	2-2-0	3	4.7
1-1-1-0	36	44.2	2-1-1-0	5	4.2
1-3-0	3	2.3	2-3-0	—	0.4
1-1-2-0	8	5.6	2-2-1	—	0.9
1-2-1-0	11	6.8	2-1-2	1	0.2
1-1-1-1-0	14	10.1	2-1-1-1	1	0.4
1-4	0	0.1			
1-3-1	0	1.1			
1-1-3	2	0.4			
1-2-2	1	1.9			
1-2-1-1	3	6.8			
1-1-2-1	2	1.6			
1-1-1-2	2	0.9			
1-1-1-1-1	4	1.5			
	664	664.2		92	92.1

Probability of transfer between any two specified members of group for

single primary = .114
double primary = .081

χ^2 (grouping up single and multiple primary outbreaks where expectations are less than 5) = 12.72
 $n = 10, .30 > P > .20$

Both the theoretical models used thus far have postulated a short period of infectiousness relative to a fairly constant incubation period. One of these

theories, which also took into account the number of infective sources present, has been shown to agree well with several aspects of the observational data. Observation is thus consistent with this hypothesis, but it gives no final proof, for the same data might be as well fitted by some alternative theory. Bailey (1955), for example, has evolved a model based on the original conception of M'Kendrick (1926) of a constant ratio of the rate of infection to the rate of removal of susceptibles from circulation by death, isolation, or recovery. Here, the period of infectiousness is not relatively instantaneous but has a distribution like the die-away negative exponential seen *e.g.* in counts of organisms in throat swabs from patients convalescing from diphtheria.

As Table VI shows, this model gives a very good fit to the total number of colds observed in completed family epidemics after a single primary. Indeed, the fit is better than predictions based on the Reed-Frost theory. Bailey himself found his model more appropriate to data collected in outbreaks of scarlet fever, where the period of infectiousness is relatively long, than to data from measles epidemics where the duration of infectiousness is very brief compared with the incubation period. In the common cold, therefore, it appears that, although the simpler theoretical model of chance contact between briefly infective individuals gives an adequate description of the pattern of spread, a more refined theory based on a prolonged period of infectivity of varying length and a consequent variable incubation period agrees better with at least one aspect of the observations.

TABLE VI
OBSERVED AND EXPECTED DISTRIBUTIONS OF SIZE
OF COMPLETED FAMILY EPIDEMICS
(BAILEY MODEL)

No. of Cases	Degree of Domestic Overcrowding					
	Overcrowded		Crowded		Uncrowded	
	Observed	Expected	Observed	Expected	Observed	Expected
0	112	113.5	155	154.2	156	166.3
1	35	32.3	41	42.9	55	42.1
2	17	18.1	24	23.3	19	19.9
3	11	11.5	15	14.0	10	10.0
4	6	5.7	6	6.6	2	3.7
Total	181	181.1	241	241.0	242	242.0
χ^2	0.35		0.24		5.41	
<i>P</i>	.95 > .90		.98 > .95		.20 > .10	

In the discussion following the paper by Bailey (1955), Armitage suggested that a simultaneous variation in the chance of the transfer of infection between individuals and between families living in

different circumstances might act in opposite directions to produce a spuriously good fit to a mathematical model which had not taken these factors into account. Table II has already shown that the individuals do vary in this respect, and Table IV confirms that environmental overcrowding also affects the risks of cross infection. Within each housing group, however, there is a reasonably close agreement between theoretical expectation based on the Reed-Frost model and actual observation.

When the Bailey continuous infection model is applied to the same housing groups, there is a disconcertingly good fit to the epidemics in the families living in overcrowded homes with either one or two rooms, or in crowded houses with three rooms. On the other hand, the fit for those living in uncrowded conditions was hardly as good as that for the brief infection period model. Without further tests of the appropriateness of each model, *e.g.* in the description of the epidemic chains as well as in the estimation of total epidemic size, it is impossible to be sure that the continuous infection model does give a better description of the epidemic pattern in a crowded household where bed isolation of the infective case is less easy; and, so far, no theory of the latter type has been suggested which can be applied to the description of epidemic chaining. Nor can it be said that the instantaneous infection model is more appropriate to the conditions in larger and less crowded houses where effective exposure times are likely to be relatively short. Nevertheless, the application of such different models to observed facts illustrates the prospect of their use in deciding the most likely mechanism of epidemic behaviour in different environmental circumstances.

SUMMARY AND CONCLUSIONS

Time charts of outbreaks of the common cold in families living in different circumstances of domestic crowding have been used to test several theoretical models of epidemic behaviour. All these models have been designed to take into account the large element of chance in the risk of effective contact between infectious case and susceptible in small family groups.

In the first generation of family outbreaks, the simplest binomial laws of chance behaviour apply quite closely to the observed distribution of numbers of cases in this secondary wave.

Epidemic theories such as that proposed by Greenwood, which assume a brief period of infectiousness followed by a relatively long and stable incubation period, do appear to fit the observed distribution of the total numbers affected in a series of family outbreaks. A better fit is given by the Reed-Frost theory,

which also postulates an increased risk of infection when more than one infectious case is circulating in the home. The fit of this latter model is further improved when it is applied within groups of outbreaks where either the status of the introducing case or the degree of domestic overcrowding is the same.

When outbreaks are grouped according to the pattern of the serial occurrence of cases, there is close agreement between observation and expectation based on the Reed-Frost theoretical model. On the other hand, an alternative theory, developed by Bailey and based on the mixing of continuously infective sources with a group of susceptibles among whom the incubation period varies, gives a very close fit particularly to data from outbreaks in crowded homes. Unfortunately, no corresponding stochastic theory is yet available for the prediction of the time sequence of cases within the epidemic.

It appears that these theoretical models of epidemic behaviour are by no means unrealistic and that they hold out prospects of fruitful application in the interpretation of data collected by field inquiry. There is a clear need for the collection of such data and for the development of alternative theories of

the mechanics of epidemic behaviour for testing against the observed facts.

This study was begun during the tenure of a Medical Research Council Grant by one of us (M.A.H.). We are grateful to Dr. Norman Bailey and Dr. Ian Sutherland for their criticisms and suggestions, and to Miss Linda Colwell and Mrs. Barbara Hunt for computing assistance.

REFERENCES

- Badger, G. F., Dingle, J. H., Feller, A. E., Hodges, R. G., Jordan, W. S., and Rammelkamp, C. H. (1953). *Amer. J. Hyg.*, **58**, 174.
 Bailey, N. T. J. (1955). *J. roy. statist. Soc.*, **B**, **17**, 35.
 — (1957). "The Mathematical Theory of Epidemics." Griffin, London.
 Brimblecombe, F. S. W., Cruickshank, R., Masters, P. L., Reid, D. D., and Stewart, G. T. (1958). *Brit. med. J.*, **1**, 119.
 Greenwood, M. (1931). *J. Hyg. (Camb.)*, **31**, 336.
 Hope Simpson, R. E., and Sutherland, I. (1954). *Lancet*, **1**, 721.
 Lidwell, O. M., and Sommerville, T. (1951). *J. Hyg. (Camb.)*, **49**, 365.
 MacDonald, G. (1953). *Trop. Dis. Bull.*, **50**, 871.
 M'Kendrick, A. G. (1926). *Proc. Edinb. math. Soc.*, **44**, 98.
 Ross, R. (1915). *Brit. med. J.*, **1**, 546.